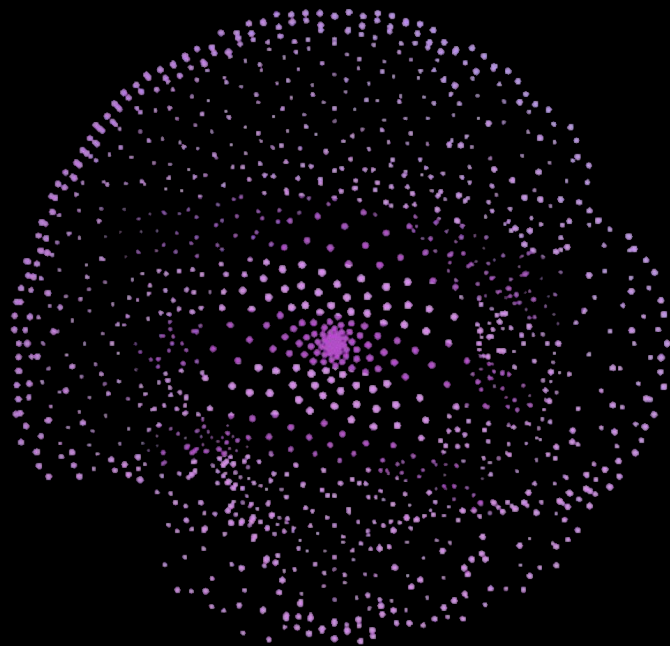




How Spotify built a SQL layer over our user listening history for agentic access



Intro



kianmehrabani



kiancm

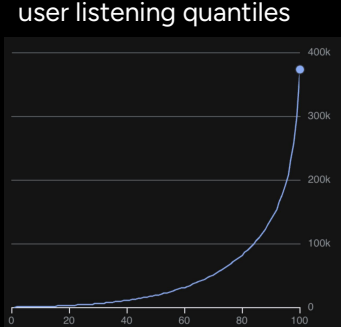


Personalization



User Behavior Data Platform

- user-keyed behavior facts
- petabyte-scale
- low-latency access for recent data
- precomputed views for long-range access
- Agents???



```
message Stream {  
    string user_id = 1;  
    string uri = 2;  
    uint64 timestamp_ms = 3;  
    ...  
}
```



Agents!

what is my top artist in 2019?

what were my most played songs last summer?

what was my first stream?

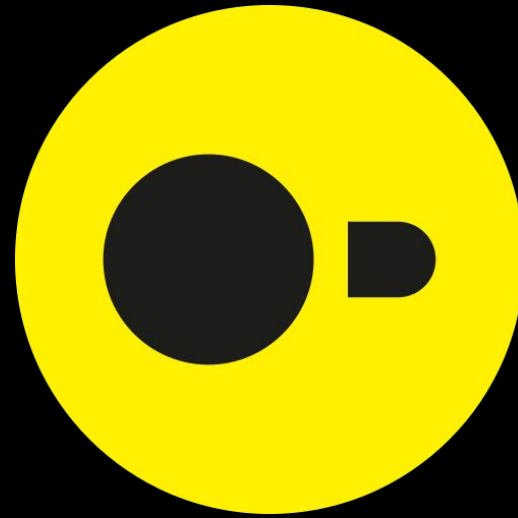
how has my music taste changed this year?

play me something I have never listened to before

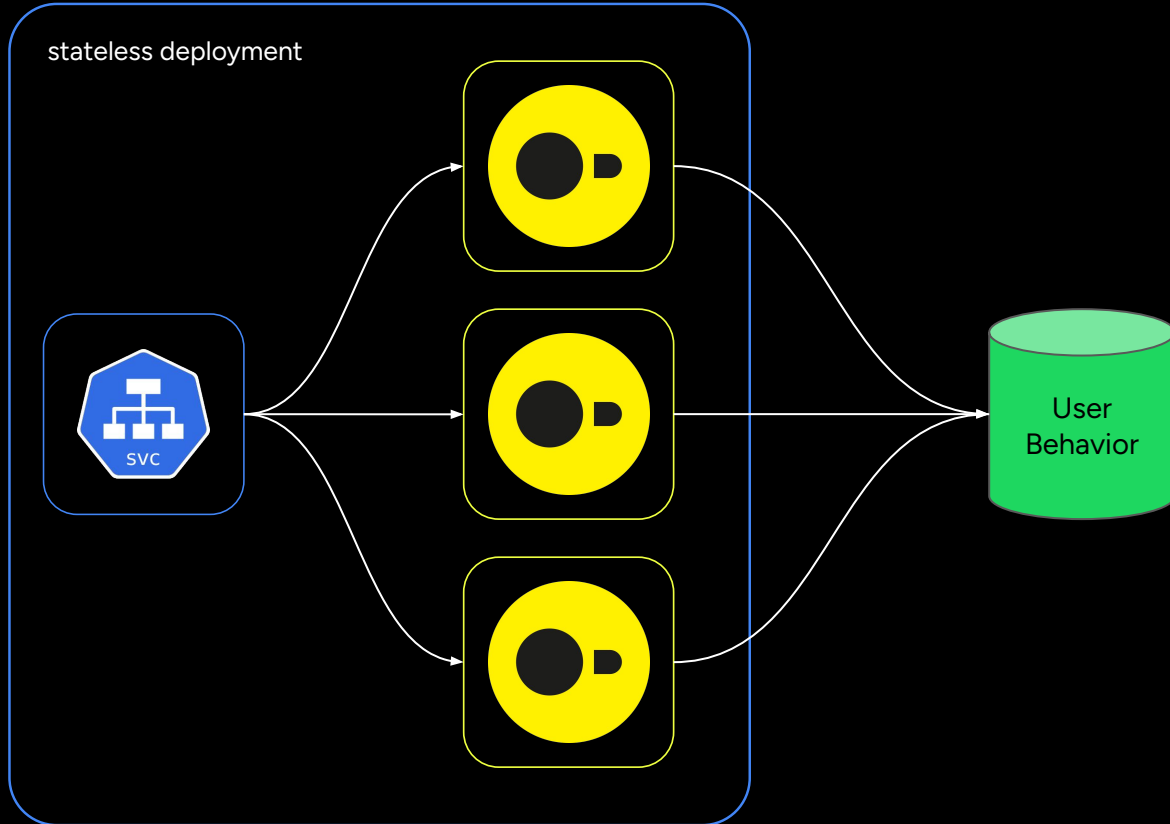
Need: flexible, dynamic query interface to user behavior



Walks and Quacks like...

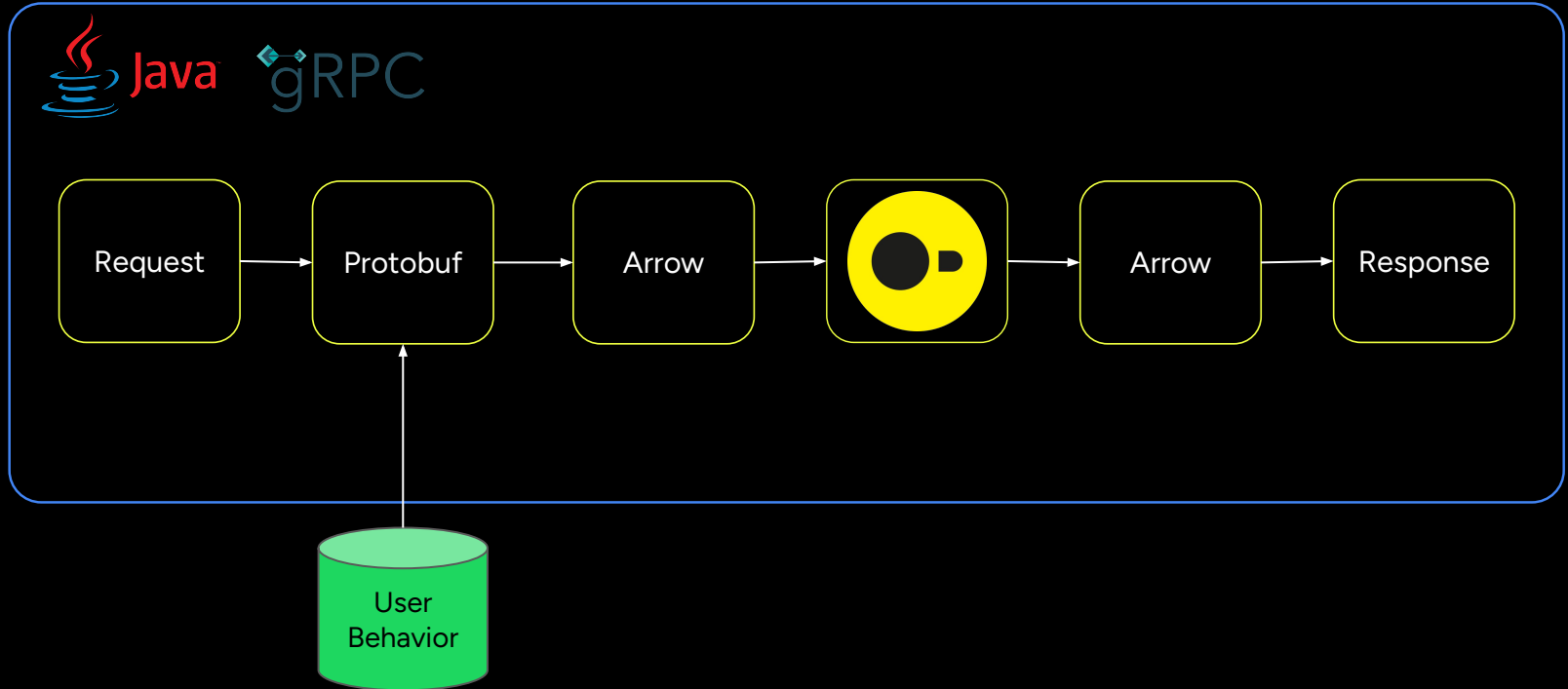


Architecture





Architecture



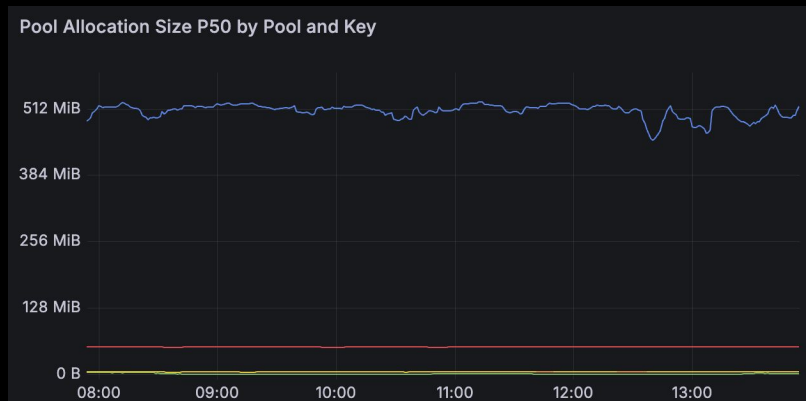
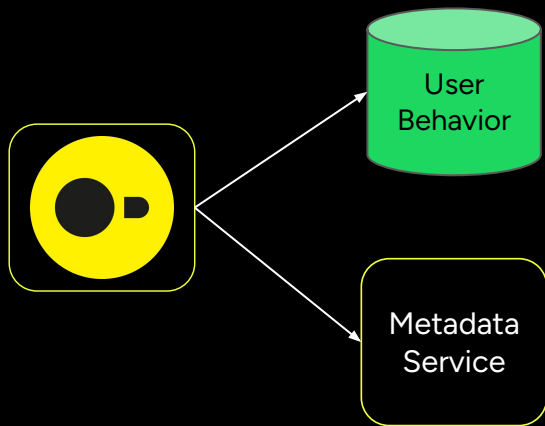
Learnings - Query Patterns

- user playcounts by content over different time windows
- first & last plays of a track
- time ranges from days to years



More Learnings

- memory management
- supporting joins
- democratizing data access



jemalloc before & after

(faucet-sql-service) Memory Usage - Avg per pod ⓘ





Questions?

